

OIPE

RAW SEQUENCE LISTING

DATE: 09/18/2001

PATENT APPLICATION: US/09/700,993

TIME: 13:15:30

Input Set : A:\81356154.app

Output Set: N:\CRF3\09182001\I700993.raw

ENTERED

p. 5

3 <110> APPLICANT: KOBAYASHI, KAZUO
 4 TAKEUCHI, MAKOTO
 5 IWAMATSU, AKIHIKO
 6 YAMAMOTO, KENJI
 7 KUMAGAI, HIDEHIKO
 8 YOSHIDA, SATOSHI
 10 <120> TITLE OF INVENTION: ENDO-BETA-N-ACETYLGLUCOSAMINIDASE GENE
 12 <130> FILE REFERENCE: 081356/0154
 14 <140> CURRENT APPLICATION NUMBER: 09/700,993
 15 <141> CURRENT FILING DATE: 2000-11-22
 17 <150> PRIOR APPLICATION NUMBER: JP98/141717
 18 <151> PRIOR FILING DATE: 1998-05-22
 20 <160> NUMBER OF SEQ ID NOS: 37
 22 <170> SOFTWARE: PatentIn Ver. 2.1
 24 <210> SEQ ID NO: 1
 25 <211> LENGTH: 2369
 26 <212> TYPE: DNA
 27 <213> ORGANISM: Mucor hiemalis
 29 <400> SEQUENCE: 1
 30 gtcgacccac gcgtccgcgcg acgcgtgggc ggacgcgtgg gcggacgcgt gggttttatt 60
 31 ttacataaat atgccttcac ttcaattgca acctgatgac aaactagcac ctgtttcttt 120
 32 tgcacttaag tctatgaatg agttgagggg ctggacgcca gacgaaaaga taaagttaa 180
 33 cgtttcaagc gtggcactac agcctcgtgt gaaaaacgcc ctgaaacctc aattattgtt 240
 34 aactcatgat atggcaggag gatataaaga agataaaaat attcaaggaa acaattataa 300
 35 agacatttat aacattcaat attggcattt agctgatact tttgtatatt tctctcatga 360
 36 gcgagtttagc attcctccag tcaattggac aaatgcttgt catagaaatg gtgtaaagt 420
 37 tttaggtact tttttagtag aaggaaataa ccaaatgcat gaaatggaag ccttgcttca 480
 38 cgggtccacct ttacttaata aactgacga ccctatgaga ttatggagtc cgtattatgc 540
 39 agaccaatta gttgctattg ctaaactacta tggttttgat ggctggttgt tcaatattga 600
 40 atgcgaattc tttccttttc ctacaaatcc aaaattcaaa gctgaagagt tggcaaagtt 660
 41 tctacactat ttttaaggaaa aattgcataa cgaaatacct ggatctcaac tcatttggtg 720
 42 cgacagcatg acaaatgaag gagaaatcca ctggcagaac cagctcacat ggaaaaatga 780
 43 gttatttttt aaaaacacgc atggtatttt tttgaattat tgggtgaaaa aagaataccc 840
 44 tgaaatggcg cgtagagtag ctgaaggaat aggtagatca ggtttagaag tttatttttg 900
 45 tacagatgta tggggaaggc atacttatgg tggcgggtgg ttcaaatcat ataagggtgt 960
 46 aaaaactgcc tactctgcaa tgacatcttc tgcattattt ggtatggcat ggacatacga 1020
 47 gcatttcgaa aagctggaat ttgaaaagat ggatcgtttg ttttggtgtg gtggtaaata 1080
 48 ctctgactat cctcccccac ctcttaaaaa ccagatgac gaaaaagaag tagaaagcga 1140
 49 tgatagtga gtagagtcga tgtacggaca caagaaaggt attgctgaca cggtagaatc 1200
 50 tattcctgta ccaggaacag attggtttgt taccaatttt gataggggtt ttggaaatag 1260
 51 gttttattat agaggaaaga gattactttc tcagccttgg tcccatttat cgcatacagc 1320
 52 tattctcccc aataaaagct atcgaaatcc agagatttat cccactgatc aaaacattaa 1380
 53 aatcactagt tctctcgatt gcgatcatgg agcttttctt ggtggaacct cgcttattat 1440
 54 caaaggccaa cgtttcaatc atagagaatc gcatgatgtt gaaactgaaa ttagtatacc 1500
 55 tctgtataag ctttcattag atgctagtaa aggatgtcga ttgcgttata tttatagaac 1560
 56 tttgttgatg aaagatgtaa agttgacagt agcatgtcac ttttcgttaa aaacaaacga 1620
 57 ctcatgtaat ttcttcaagg tatggcagcc agatgaaaat ttctcttttg aatatgatga 1680

RAW SEQUENCE LISTING

DATE: 09/18/2001

PATENT APPLICATION: US/09/700,993

TIME: 13:15:30

Input Set : A:\81356154.app

Output Set: N:\CRF3\09182001\I700993.raw

```

58 tggaaatgaga gccactgtta caactgaaaa ttctaccgaa agcagatgct ttttattacg 1740
59 tacaacagaa gaagatacag gagaaaatga ttggataaca aaaactatta atgtgcctgc 1800
60 tgttccagaa ggaagtcaat tatacattac aagacttgaa gtgagcgtag tattagatac 1860
61 agctgggttta gtaggtccttg ttaatcaagt tattgcttgc ttgggatata ttagcatcat 1920
62 accaactata aattctggaa taaaaacaga ttcatacgcg attattcagg atctcttttg 1980
63 gaaagatcag aaatatacca aaatcggaag agaaagttta gacgacatag ctcaagaaga 2040
64 agttcataga tattatggaa cattgaactg ggaaaacaca gcaaagttag taaacgcttg 2100
65 ggaggaaata gattactaca acgtttttta caaagaaagt gacgactctg caactcgcac 2160
66 ctttttagga acagcattct gtaatcaatt tcgtgtatct ggttttagata ttattttatc 2220
67 taagctacca aagatagtta ttgaagctgt taacaaagaa ggatacatct cttcaagtgg 2280
68 tagcatagat ttgtcattaa actaggactt gaaataaaat attatgataa agaaaaaaa 2340
69 aaaaaaaaaa aaaaaaaaaa ggcggccgc 2369
72 <210> SEQ ID NO: 2
73 <211> LENGTH: 2235
74 <212> TYPE: DNA
75 <213> ORGANISM: Mucor hiemalis
77 <220> FEATURE:
78 <221> NAME/KEY: CDS
79 <222> LOCATION: (1)..(2232)
81 <400> SEQUENCE: 2
82 atg cct tca ctt caa ttg caa cct gat gac aaa cta gca cct gtt tct 48
83 Met Pro Ser Leu Gln Leu Gln Pro Asp Asp Lys Leu Ala Pro Val Ser
84 1 5 10 15
86 ttt gca ctt aag tct atg aat gag ttg agg gac tgg acg cca gac gaa 96
87 Phe Ala Leu Lys Ser Met Asn Glu Leu Arg Asp Trp Thr Pro Asp Glu
88 20 25 30
90 aag ata aag ttt aac gtt tca agc gtg gca cta cag cct cgt gtg aaa 144
91 Lys Ile Lys Phe Asn Val Ser Ser Val Ala Leu Gln Pro Arg Val Lys
92 35 40 45
94 aac gcc ctg aaa cct caa tta ttg tta act cat gat atg gca gga gga 192
95 Asn Ala Leu Lys Pro Gln Leu Leu Leu Thr His Asp Met Ala Gly Gly
96 50 55 60
98 tat aaa gaa gat aaa aat att caa gga aac aat tat aaa gac att tat 240
99 Tyr Lys Glu Asp Lys Asn Ile Gln Gly Asn Asn Tyr Lys Asp Ile Tyr
100 65 70 75 80
102 aac att caa tat tgg cat tta gct gat act ttt gta tat ttc tct cat 288
103 Asn Ile Gln Tyr Trp His Leu Ala Asp Thr Phe Val Tyr Phe Ser His
104 85 90 95
106 gag cga gtt agc att cct cca gtc aat tgg aca aat gct tgt cat aga 336
107 Glu Arg Val Ser Ile Pro Pro Val Asn Trp Thr Asn Ala Cys His Arg
108 100 105 110
110 aat ggt gta aag tgt tta ggt act ttt tta gta gaa gga aat aac caa 384
111 Asn Gly Val Lys Cys Leu Gly Thr Phe Leu Val Glu Gly Asn Asn Gln
112 115 120 125
114 atg cat gaa atg gaa gcc ttg ctt cac ggt cca cct tta ctt aat aac 432
115 Met His Glu Met Glu Ala Leu Leu His Gly Pro Pro Leu Leu Asn Asn
116 130 135 140
118 act gac gac cct atg aga tta tgg agt ccg tat tat gca gac caa tta 480
119 Thr Asp Asp Pro Met Arg Leu Trp Ser Pro Tyr Tyr Ala Asp Gln Leu

```

RAW SEQUENCE LISTING

DATE: 09/18/2001

PATENT APPLICATION: US/09/700,993

TIME: 13:15:30

Input Set : A:\81356154.app

Output Set: N:\CRF3\09182001\I700993.raw

120	145				150				155				160				
122	ggt	gct	att	gct	aaa	cac	tat	ggt	ttt	gat	ggc	tgg	ttg	ttc	aat	att	528
123	Val	Ala	Ile	Ala	Lys	His	Tyr	Gly	Phe	Asp	Gly	Trp	Leu	Phe	Asn	Ile	
124					165				170				175				
126	gaa	tgc	gaa	ttc	ttt	cct	ttt	cct	aca	aat	cca	aaa	ttc	aaa	gct	gaa	576
127	Glu	Cys	Glu	Phe	Phe	Pro	Phe	Pro	Thr	Asn	Pro	Lys	Phe	Lys	Ala	Glu	
128				180					185				190				
130	gag	ttg	gca	aag	ttt	cta	cac	tat	ttt	aag	gaa	aaa	ttg	cat	aac	gaa	624
131	Glu	Leu	Ala	Lys	Phe	Leu	His	Tyr	Phe	Lys	Glu	Lys	Leu	His	Asn	Glu	
132			195					200					205				
134	ata	cct	gga	tct	caa	ctc	att	tgg	tac	gac	agc	atg	aca	aat	gaa	gga	672
135	Ile	Pro	Gly	Ser	Gln	Leu	Ile	Trp	Tyr	Asp	Ser	Met	Thr	Asn	Glu	Gly	
136		210				215						220					
138	gaa	atc	cac	tgg	cag	aac	cag	ctc	aca	tgg	aaa	aat	gag	tta	ttt	ttt	720
139	Glu	Ile	His	Trp	Gln	Asn	Gln	Leu	Thr	Trp	Lys	Asn	Glu	Leu	Phe	Phe	
140	225				230					235				240			
142	aaa	aac	acg	gat	ggt	att	ttt	ttg	aat	tat	tgg	tgg	aaa	aaa	gaa	tac	768
143	Lys	Asn	Thr	Asp	Gly	Ile	Phe	Leu	Asn	Tyr	Trp	Trp	Lys	Lys	Glu	Tyr	
144			245					250					255				
145	cct	gaa	atg	gcg	cgt	aga	gta	gct	gaa	gga	ata	ggt	aga	tca	ggt	tta	816
146	Pro	Glu	Met	Ala	Arg	Arg	Val	Ala	Glu	Gly	Ile	Gly	Arg	Ser	Gly	Leu	
147			260					265					270				
149	gaa	gtt	tat	ttt	ggt	aca	gat	gta	tgg	gga	agg	cat	act	tat	ggt	ggc	864
150	Glu	Val	Tyr	Phe	Gly	Thr	Asp	Val	Trp	Gly	Arg	His	Thr	Tyr	Gly	Gly	
151			275				280					285					
153	ggt	ggt	ttc	aaa	tca	tat	aag	ggt	gta	aaa	act	gcc	tac	tct	gca	atg	912
154	Gly	Gly	Phe	Lys	Ser	Tyr	Lys	Gly	Val	Lys	Thr	Ala	Tyr	Ser	Ala	Met	
155		290				295					300						
157	aca	tct	tct	gca	tta	ttt	ggt	atg	gca	tgg	aca	tac	gag	cat	ttc	gaa	960
158	Thr	Ser	Ser	Ala	Leu	Phe	Gly	Met	Ala	Trp	Thr	Tyr	Glu	His	Phe	Glu	
159	305				310					315					320		
161	aag	tct	gaa	ttt	gaa	aag	atg	gat	cgt	ttg	ttt	tgg	tgt	ggt	ggt	aaa	1008
162	Lys	Ser	Glu	Phe	Glu	Lys	Met	Asp	Arg	Leu	Phe	Trp	Cys	Gly	Gly	Lys	
163				325				330					335				
165	tac	tct	gac	tat	cct	ccc	cca	cct	cct	aaa	aac	cca	gat	gac	gaa	aaa	1056
166	Tyr	Ser	Asp	Tyr	Pro	Pro	Pro	Pro	Pro	Lys	Asn	Pro	Asp	Asp	Glu	Lys	
167			340					345					350				
169	gaa	gta	gaa	agc	gat	gat	agt	gaa	gat	gag	ctc	atg	tac	gga	cac	aag	1104
170	Glu	Val	Glu	Ser	Asp	Asp	Ser	Glu	Asp	Glu	Leu	Met	Tyr	Gly	His	Lys	
171			355				360					365					
173	aaa	ggt	att	gct	gac	acg	gta	gaa	tct	att	cct	gta	cca	gga	aca	gat	1152
174	Lys	Gly	Ile	Ala	Asp	Thr	Val	Glu	Ser	Ile	Pro	Val	Pro	Gly	Thr	Asp	
175		370				375					380						
177	tgg	ttt	gtt	acc	aat	ttt	gat	agg	ggg	ttt	gga	aat	agg	ttt	tat	tat	1200
178	Trp	Phe	Val	Thr	Asn	Phe	Asp	Arg	Gly	Phe	Gly	Asn	Arg	Phe	Tyr	Tyr	
179	385				390				395					400			
181	aga	gga	aag	aga	tta	ctt	tct	cag	cct	tgg	tcc	cat	tta	tcg	cat	caa	1248
182	Arg	Gly	Lys	Arg	Leu	Leu	Ser	Gln	Pro	Trp	Ser	His	Leu	Ser	His	Gln	
183				405				410					415				

RAW SEQUENCE LISTING

DATE: 09/18/2001

PATENT APPLICATION: US/09/700,993

TIME: 13:15:30

Input Set : A:\81356154.app

Output Set: N:\CRF3\09182001\I700993.raw

```

185 gct att ctc ccc aat aaa agc tat cga aat cca gag att tat ccc act 1296
186 Ala Ile Leu Pro Asn Lys Ser Tyr Arg Asn Pro Glu Ile Tyr Pro Thr
187          420          425          430
189 gat caa aac att aaa atc act agt tct ctc gat tgc gat cat gga gct 1344
190 Asp Gln Asn Ile Lys Ile Thr Ser Ser Leu Asp Cys Asp His Gly Ala
191          435          440          445
193 ttt ctt ggt gga acc tcg ctt att atc aaa ggc caa cgt ttc aat cat 1392
194 Phe Leu Gly Gly Thr Ser Leu Ile Ile Lys Gly Gln Arg Phe Asn His
195          450          455          460
197 aga gaa tcg cat gat gtt gaa act gaa att agt ata cct ctg tat aag 1440
198 Arg Glu Ser His Asp Val Glu Thr Glu Ile Ser Ile Pro Leu Tyr Lys
199 465          470          475          480
201 ctt tca tta gat gct agt aaa gga tgc tca ttg cgt tat att tat aga 1488
202 Leu Ser Leu Asp Ala Ser Lys Gly Cys Ser Leu Arg Tyr Ile Tyr Arg
203          485          490          495
205 act ttg ttg atg aaa gat gta aag ttg aca gta gca tgt cac ttt tcg 1536
206 Thr Leu Leu Met Lys Asp Val Lys Leu Thr Val Ala Cys His Phe Ser
207          500          505          510
209 tta aaa aca aac gac tca gtt aat ttc ttc aag gta tgg cag cca gat 1584
210 Leu Lys Thr Asn Asp Ser Val Asn Phe Phe Lys Val Trp Gln Pro Asp
211          515          520          525
213 gaa aat ttc tct ttt gaa tat gat gat gga atg aga gcc act gtt aca 1632
214 Glu Asn Phe Ser Phe Glu Tyr Asp Asp Gly Met Arg Ala Thr Val Thr
215          530          535          540
217 act gaa aat tct acc gaa agc aga tgc ttt tta tta cgt aca aca gaa 1680
218 Thr Glu Asn Ser Thr Glu Ser Arg Cys Phe Leu Leu Arg Thr Thr Glu
219 545          550          555          560
221 gaa gat aca gga gaa aat gat tgg ata aca aaa act att aat gtg cct 1728
222 Glu Asp Thr Gly Glu Asn Asp Trp Ile Thr Lys Thr Ile Asn Val Pro
223          565          570          575
225 gct gtt cca gaa gga agt caa tta tac att aca aga ctt gaa gtg agc 1776
226 Ala Val Pro Glu Gly Ser Gln Leu Tyr Ile Thr Arg Leu Glu Val Ser
227          580          585          590
229 gta gta tta gat aca gct ggt tta gta ggt ctt gtt aat caa gtt att 1824
230 Val Val Leu Asp Thr Ala Gly Leu Val Gly Leu Val Asn Gln Val Ile
231          595          600          605
233 gct tgc ttg gga tat att agc atc ata cca act ata aat tct gga ata 1872
234 Ala Cys Leu Gly Tyr Ile Ser Ile Ile Pro Thr Ile Asn Ser Gly Ile
235          610          615          620
237 aaa aca gat tca tca cgc att att cag gat ctc ttt tgg aaa gat cag 1920
238 Lys Thr Asp Ser Ser Arg Ile Ile Gln Asp Leu Phe Trp Lys Asp Gln
239 625          630          635          640
241 aaa tat acc aaa atc gga aaa gaa agt tta gac gac ata gct caa gaa 1968
242 Lys Tyr Thr Lys Ile Gly Lys Glu Ser Leu Asp Asp Ile Ala Gln Glu
243          645          650          655
245 gaa gtt cat aga tat tat gga aca ttg aac tgg gaa aac aca gca aat 2016
246 Glu Val His Arg Tyr Tyr Gly Thr Leu Asn Trp Glu Asn Thr Ala Asn
247          660          665          670
249 gta gta aac gct tgg gag gaa ata gat tac tac aac gtt ttt tac aaa 2064

```

RAW SEQUENCE LISTING

DATE: 09/18/2001

PATENT APPLICATION: US/09/700,993

TIME: 13:15:30

Input Set : A:\81356154.app

Output Set: N:\CRF3\09182001\I700993.raw

```

250 Val Val Asn Ala Trp Glu Glu Ile Asp Tyr Tyr Asn Val Phe Tyr Lys
251      675      680      685
253 gaa agt gac gac tct gca act cgc atc ttt tta gga aca gca ttc tgt 2112
254 Glu Ser Asp Asp Ser Ala Thr Arg Ile Phe Leu Gly Thr Ala Phe Cys
255      690      695      700
257 aat caa ttt cgt gta tct ggt tta gat att att tta tct aag cta cca 2160
258 Asn Gln Phe Arg Val Ser Gly Leu Asp Ile Ile Leu Ser Lys Leu Pro
259 705      710      715      720
261 aag ata gtt att gaa gct gtt aac aaa gaa gga tac atc tct tca agt 2208
262 Lys Ile Val Ile Glu Ala Val Asn Lys Glu Gly Tyr Ile Ser Ser Ser
263      725      730      735
265 ggt agc ata gat ttg tca tta aac tag 2235
266 Gly Ser Ile Asp Leu Ser Leu Asn
267      740
270 <210> SEQ ID NO: 3
271 <211> LENGTH: 744
272 <212> TYPE: PRT
273 <213> ORGANISM: Mucor hiemalis
275 <400> SEQUENCE: 3
276 Met Pro Ser Leu Gln Leu Gln Pro Asp Asp Lys Leu Ala Pro Val Ser
277 1      5      10      15
279 Phe Ala Leu Lys Ser Met Asn Glu Leu Arg Asp Trp Thr Pro Asp Glu
280      20      25      30
282 Lys Ile Lys Phe Asn Val Ser Ser Val Ala Leu Gln Pro Arg Val Lys
283      35      40      45
285 Asn Ala Leu Lys Pro Gln Leu Leu Leu Thr His Asp Met Ala Gly Gly
286      50      55      60
288 Tyr Lys Glu Asp Lys Asn Ile Gln Gly Asn Asn Tyr Lys Asp Ile Tyr
289 65      70      75      80
291 Asn Ile Gln Tyr Trp His Leu Ala Asp Thr Phe Val Tyr Phe Ser His
292      85      90      95
294 Glu Arg Val Ser Ile Pro Pro Val Asn Trp Thr Asn Ala Cys His Arg
295      100      105      110
297 Asn Gly Val Lys Cys Leu Gly Thr Phe Leu Val Glu Gly Asn Asn Gln
298      115      120      125
299 Met His Glu Met Glu Ala Leu Leu His Gly Pro Pro Leu Leu Asn Asn
300      130      135      140
302 Thr Asp Asp Pro Met Arg Leu Trp Ser Pro Tyr Tyr Ala Asp Gln Leu
303 145      150      155      160
305 Val Ala Ile Ala Lys His Tyr Gly Phe Asp Gly Trp Leu Phe Asn Ile
306      165      170      175
308 Glu Cys Glu Phe Phe Pro Phe Pro Thr Asn Pro Lys Phe Lys Ala Glu
309      180      185      190
311 Glu Leu Ala Lys Phe Leu His Tyr Phe Lys Glu Lys Leu His Asn Glu
312      195      200      205
314 Ile Pro Gly Ser Gln Leu Ile Trp Tyr Asp Ser Met Thr Asn Glu Gly
315      210      215      220
317 Glu Ile His Trp Gln Asn Gln Leu Thr Trp Lys Asn Glu Leu Phe Phe
318 225      230      235      240

```

Use of n and / or Xaa has been detected in the
Sequence Listing. Review the Sequence Listing
to ensure a corresponding explanation is present
in the <220> to <223> fields of each sequence
using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/700,993

DATE: 09/18/2001

TIME: 13:15:31

Input Set : A:\81356154.app

Output Set: N:\CRF3\09182001\I700993.raw

L:447 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:480 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7
L:499 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8
L:655 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:20
L:705 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:21